



Le Traitement Automatique des Langues au service du vin

Antoine Gérard, Benoît Sagot, Emilie Pons

► To cite this version:

Antoine Gérard, Benoît Sagot, Emilie Pons. Le Traitement Automatique des Langues au service du vin. Dataquitaine 2021 - IA, Recherche Opérationnelle & Data Science, Feb 2021, Bordeaux / Virtual, France. hal-03146219

HAL Id: hal-03146219

<https://inria.hal.science/hal-03146219>

Submitted on 18 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le Traitement Automatique des Langues au service du vin

Gérard Antoine, antoine.gerard@winespace.fr, Winespace, Bordeaux (Orateur)

Sagot Benoît, Inria, Paris, benoit.sagot@inria.fr

Inria Bordeaux, stip-bso@inria.fr

Thème : Intelligence Artificielle

Résumé : Dans cette présentation, nous proposons de détailler une collaboration fructueuse entre l'institut de recherche Inria et une startup bordelaise : Winespace. Nous nous intéresserons alors à l'analyse sémantique de commentaires de dégustation dans le but de recommander des vins présentant des caractéristiques similaires.

Mots clés : *Traitement automatique des Langues, Analyse sémantique, Œnologie*

1. Introduction

Winespace souhaitait mener à bien un projet de recommandation de vins similaires basé sur l'étude de commentaires de dégustation. Pour cela, une collaboration entre Winespace et Inria a été mise en place. Côté Inria, elle impliquait les centres de recherche de Bordeaux et Paris, en particulier les équipes Transfert, Innovation et Partenariat et Service Expérimentation et Développement de Bordeaux et l'équipe ALMAAnaCH de Paris.

2. Méthodologie

Pour ce projet, nous avons mis en place des outils relevant du Traitement Automatique des Langues (TAL) appliqués au domaine du vin. Nous avons alors réalisé un algorithme de compréhension automatique des commentaires de dégustation, dans le but de permettre la mise en correspondance de vins dont les commentaires montrent qu'ils ont des caractéristiques communes, notamment à des fins de recommandation.

Deux difficultés émanent alors de ce projet. La première est le caractère hétérogène des commentaires de dégustation, dont le contenu, le vocabulaire et le style diffèrent d'un œnologue à l'autre. La deuxième difficulté est l'absence de données annotées : aucun commentaire d'œnologue n'a été manuellement annoté pour y repérer les termes pertinents ou l'associer à des informations structurées. Réaliser de telles annotations serait coûteux, et le nombre élevé de termes rares rendrait la tâche ardue et le volume d'annotation à produire très élevé.

En revanche, la disponibilité d'une première version d'ontologie du domaine, fournie par Winespace, permet de déployer des technologies symboliques et statistiques ne reposant pas sur l'apprentissage automatique. C'est donc ce choix que nous avons effectué. Il a pour autre avantage de fournir des modèles analysables, dont les résultats sont donc explicables, mais également améliorables (par exemple par l'amélioration progressive de l'ontologie).

3. Originalité / perspective

Ce projet avait pour objectif la construction d'un système de recommandations de vins similaires basé sur les commentaires de dégustation en entrée. Après avoir réalisé une

analyse sémantique des commentaires, chaque vin a été représenté par un vecteur « profil » permettant d'encoder tous les concepts présents dans celui-ci : arômes, acidité, amertume, ...

L'algorithme mis en place est alors une preuve de concept prometteuse. L'amélioration de l'ontologie fournie par Winespace, ainsi que l'utilisation de nouvelles techniques d'analyse sémantique permettront sûrement de perfectionner cet outil, et ainsi permettre à Winespace d'analyser plus facilement les commentaires de dégustation.

Cette preuve de concept ouvre également des perspectives prometteuses sur une meilleure compréhension du goût du vin.